Please amend the paragraph of page 28, lines 23 to 29, as follows:

-- The second test of the interactions predicted by the Rosetta Stone method uses as confirmation the Database of Interacting Proteins provided at the website of the UCLA DOE laboratory. This is a compilation of protein pairs that have been found to interact in some published experiment. As of December 1998, the database contained 939 entries, 724 of which have both members of the pair listed in the ProDom database. Of these 724 pairs, we find 46 or 6.4% linked by Rosetta Stone sequences. We expect this percentage to rise as more genomes are sequenced, revealing more linked sequences. --

In The Claims:

Please cancel claims 1 and 2, without prejudice.

Please add the following new claims:

--3. (NEW) A method for identifying a high confidence functional link between at least two proteins, comprising the following steps:

(a) identifying non-homologous proteins as being functionally linked by a "Rosetta Stone" method comprising the following steps

(i) providing amino acid sequences of a first protein and a second protein, wherein the first and second proteins are not homologous,

(ii) providing an amino acid sequence of a third protein,

(iii) aligning amino acid sequence segments from the first protein and the second protein to the amino acid sequence of the third protein, wherein the amino acid sequence segments from the first and the second protein do not align to each other with any significant sequence similarity, and

(iv) establishing whether the first and second proteins are functionally linked by determining whether a significant sequence similarity is present between the aligned amino acid sequences of step (iii), thereby identifying non-homologous proteins as being functionally linked;

(b) identifying pairs of proteins in a genome as being functionally linked by a "phylogenetic profile" method comprising the following steps

New claims

(i) providing a first plurality of protein sequences comprising substantially all protein sequences encoded by a first genome,

(ii) providing a second plurality of protein sequences comprising substantially all protein sequences encoded by one or more additional genomes,

(iii) comparing each protein sequence in the first plurality of protein sequences with substantially all the protein sequences of the second plurality of protein sequences to determine if a protein sequence in the first genome has a homolog in the one or more additional genomes based on the degree of similarity of the sequences being compared,

(iv) generating a phylogenetic profile for each protein of the first genome, wherein the phylogenetic profile is a vector or pattern whose elements indicate whether a homolog of the corresponding protein is present or absent in the one or more additional genomes, and

(v) grouping together proteins having similar phylogenetic profiles, wherein a similar phylogenetic profile indicates a functional link between the proteins; and

(c) identifying pairs of proteins that are linked in both (a) and (b), thereby identifying a high confidence functional link between at least two proteins.

4. (NEW) The method of claim 3, further comprising:

generating an expression profile for each protein of the genome where the expression profile is a vector or a pattern whose elements indicate the level of mRNA expression of the corresponding gene in two or more DNA chip experiments; and

grouping together genes having similar expression profiles where a similar expression profile indicates a functional link between proteins.

5. (NEW) The method of claim 4, further comprising displaying the functional links as networks of related proteins, comprising:

placing a plurality of proteins in a diagram such that functionally linked proteins are closer together than all other proteins; and

identifying groups of proteins that fall in a cluster in said diagram as functionally related.

6. (NEW) The method of claim 5, wherein the placing of the plurality of proteins in a diagram utilizes a computer.

7. (NEW) The method of claim 3, further comprising:

identifying functional links for a plurality of protein pairs;

placing substantially all protein pairs that are identified as functionally linked in a diagram such that functionally linked proteins are closer together than other proteins; and

identifying groups of proteins that fall in a cluster in said diagram as functionally related.

8. (NEW) The method of claim 7, wherein the placing of substantially all protein pairs in a diagram utilizes a computer.

9. (NEW) A method for identifying a high confidence functional link between at least two proteins, comprising the following steps:

(a) identifying non-homologous proteins as being functionally linked by a "Rosetta Stone" method comprising the following steps

(i) providing a pair of non-homologous protein amino acid sequences;

(ii) providing an amino acid sequence of a third protein;

(iii) aligning amino acid sequence segments from the first protein in (i) and the second protein in (i) to the amino acid sequence of (ii) where the first and the second protein of (i) are not homologues; and

(iv) establishing whether a significant sequence similarity is present between the alignments of (iii), wherein identification of a significant sequence similarity between each of the two non-homologous amino acid sequence segments from two proteins of (i) to different sequence segments of the protein of (ii) identifies the pair of proteins of (i) as being functionally linked each other;

(b) identifying pairs of proteins in a genome as being functionally linked by a "phylogenetic profile" method comprising the following steps

(i) providing a first plurality of protein sequences comprising substantially all protein sequences encoded by a first genome,

(ii) providing a second plurality of protein sequences comprising substantially all protein sequences encoded by one or more additional genomes,

(iii) comparing each protein sequence in the first plurality of protein sequences with substantially all the protein sequences of the second plurality of protein sequences to determine if a protein sequence in the first genome has a homolog in the one or more additional genomes based on the degree of similarity of the sequences being compared,

(iv) generating a phylogenetic profile for each protein of the first genome, wherein the phylogenetic profile is a vector or pattern whose elements indicate whether a homolog of the corresponding protein is present or absent in the one or more additional genomes, and

(v) grouping together proteins having similar phylogenetic profiles, wherein a similar phylogenetic profile indicates a functional link between the proteins; and

(c) identifying pairs of proteins that are linked in both (a) and (b), thereby identifying a high confidence functional link between at least two proteins.

10. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method establishing that the pair of non-homologous amino acid sequence segments of (i) have significant sequence similarities to different sequence segments of the protein of (ii) comprises showing that a computed probability (p) value is below a statistically significant threshold.

11. (NEW) The method of claim 10, wherein the probability threshold is set with respect to a value 1/N, wherein N is an integer based on the total number of protein sequences in a database.

12. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method the non-homologous amino acid sequence segments from different protein sequences of (i) are at least about 50 amino acid residues long.

13. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method the non-homologous amino acid sequence segments from different polypeptide sequences of (i) are between about 50 and about 1000 amino acid residues long.

14. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method statistically insignificant Rosetta stone links are filtered out when either protein in (i) has a plurality of homologs.

15. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method the plurality of homologues is more than about 100 homologues.

16. (NEW) The method of claim 9, wherein in the "Rosetta Stone" method statistically insignificant Rosetta Stone links are filtered out when either protein in (i) forms a plurality of Rosetta Stone links to other distinct proteins.

17. (NEW) The method of claim 16, wherein the plurality of Rosetta Stone links is more than about 100.

18. (NEW) The method of claim 17, wherein the plurality of Rosetta Stone links is more than about 25.

19. (NEW) A method for identifying a high confidence functional link between at least two proteins, comprising the following steps:

(a) identifying non-homologous proteins as being functionally linked by a "Rosetta Stone" method comprising the following steps

(i) providing amino acid sequences of a first protein and a second protein, wherein the first and second proteins are not homologous,

(ii) providing an amino acid sequence of a third protein,

(iii) aligning amino acid sequence segments from the first protein and the second protein to the amino acid sequence of the third protein, wherein the amino acid sequence

segments from the first and the second protein do not align to each other with any significant sequence similarity, and

(iv) establishing whether the first and second proteins are functionally linked by determining whether a significant sequence similarity is present between the aligned amino acid sequences of step (iii), thereby identifying non-homologous proteins as being functionally linked;

(b) identifying pairs of proteins in a genome as being functionally linked by a "phylogenetic profile" method comprising the following steps

(i) providing a first plurality of protein sequences comprising substantially all protein sequences encoded by a first genome, or, a plurality of nucleic acid sequences comprising substantially all protein-encoding nucleic acid sequences in a first genome;

(ii) providing a second plurality of protein sequences comprising substantially all protein sequences encoded by one or more additional genomes, or, a second plurality of nucleic acid sequences comprising substantially all protein-encoding nucleic acid sequences of one or more additional genomes;

(iii) comparing each protein sequence or nucleic acid sequence in the first plurality of protein sequences or nucleic acid sequences respectively with substantially all the protein sequences or nucleic acid sequences of the second plurality of protein sequences or nucleic acid sequences to determine if the protein sequence or nucleic acid sequence in the first genome has a homolog in the one or more additional genomes based on the degree of similarity of the sequences being compared;

(iv) generating a phylogenetic profile for each protein of the first genome, wherein the phylogenetic profile is a vector or pattern whose elements indicate whether a homolog of the corresponding protein or nucleic acid is present or absent in the one or more additional genomes; and

(v) grouping together proteins having similar phylogenetic profiles, wherein proteins with similar profiles are identified as being functionally linked; and

(c) identifying pairs of proteins that are linked in both (a) and (b), thereby identifying a high confidence functional link between at least two proteins.

20. (NEW) The method of claim 19, wherein the phylogenetic profile is generated using a bit type profiling method.

21. (NEW) The method of claim 19, wherein the phylogenetic profile is generated using an evolutionary distance method.

22. (NEW) The method of claim 19, wherein the phylogenetic profile is generated in a binary code describing the presence or absence of a given protein in an organism.

23. (NEW) The method of claim 19, wherein the phylogenetic profile is generated in a continuous code that describes how similar the related sequences are in the different genomes.

24. (NEW) The method of claim 19, wherein the phylogenetic profile is generated using an evolution probability process, wherein the process comprises

(a) constructing a conditional probability matrix: p(aa → aa'), where aa and aa' are any amino acids, and the conditional probability matrix is constructed by converting an amino acid substitution matrix from a log odds matrix to a conditional probability matrix;

(b) accounting for an observed alignment of the constructed conditional probability matrix by taking the product and the conditional probabilities for each aligned pair of amino acids during the alignment of the two protein sequences, represented by

$$P(p)=\prod_{n} p(aa_n \rightarrow aa'_n)\text{; and}$$

(c) determining an evolutionary distance α from powers equation:

$p'=p^{\alpha}$ (aa → aa'), maximizing for P.

25. (NEW) The method of claim 24, wherein the conditional probability matrix is defined by a Markov process with substitution rates over a fixed time interval.

26. (NEW) The method of claim 24, wherein the conversion from an amino acid substitution log odds matrix to a conditional probability matrix is represented by:

$P_\beta(i \rightarrow j) = P(j)2^{\wedge} [BLOSUM62\ ij / 2]$,

where BLOSUM62 is an amino acid substitution log odds matrix, and P(i→j) is the probability that amino acid i is replaced by amino acid j through point mutations according to BLOSUM62 scores.

27. (NEW) The method of claim 26, wherein Pj's are the abundances of amino acid j and are computed by solving a plurality of linear equations given by the normalization condition that $\sum_i P_B(i \rightarrow j) = 1$.--